Fusing Animal Biometrics with Autonomous Robotics: Drone-based Search and Individual ID of Friesian Cattle (Extended Abstract)

William Andrew Colin Greatwood Tilo Burghardt
University of Bristol
{will.andrew, colin.greatwood, tb2935}@bristol.ac.uk

Abstract

This work covers the robotic drone integration of a re-identification system for Friesian Cattle. We have built a computationally-enhanced M100 UAV platform with an on-board deep learning inference system for integrated computer vision and navigation able to autonomously find and visually identify by coat pattern individual Holstein Friesian cattle in freely moving herds. For autonomous drone-based identification we describe an approach that utilises three deep convolutional neural network architectures running live onboard the aircraft; that is, a YoloV2-based species detector, a dual-stream Convolutional Neural Network (CNN) delivering exploratory agency and an InceptionV3-based biometric Long-term Recurrent Convoluational Network (LRCN) for individual animal identification. We evaluate the performance of components offline, and also online via real-world field tests of autonomous low-altitude flight in a farm environment. The presented proof-of-concept system is a successful step towards autonomous biometric identification of individual animals from the air in open pasture environments and inside farms for tag-less AI support in farming and ecology. The work is published in full in IROS 2019 [4].

1. Introduction

This paper presents an unmanned aerial vehicle (UAV) platform with onboard deep learning inference (see Fig. 1) that autonomously locates and visually identifies individual Holstein Friesian cattle by their uniquely coloured coats in low altitude flight (approx. 10m) within a geo-fenced farm area. The task encompasses the integrated performance of species detection, exploratory agency, and individual animal identification (ID). All tasks are performed entirely onboard a custom DJI M100 quadrotor with limited computational resources, battery lifetime, and payload size. In doing so, this work attempts to assist agricultural monitoring in performing minimally-invasive cattle localisation and identification in the field. Possible applications include the behavioural analysis of social hierarchies [46, 19, 35] grazing patterns [13, 14] and herd welfare [43]. The search for targets with unknown locations traditionally arises in search and rescue (SAR) scenarios [25, 47, 40]. In visually-supported navigation for this task, approaches broadly operate either a map-based or map-less paradigm [6, 9]. Map-less approaches have no global environment representation and traditionally operate using template appearance matching [29, 18], optical-flow guidance [42], or landmark feature tracking [42, 41]. More recently, such systems have been replaced with visual input classification via CNNs [12, 37]. In this work, we build on a simulation setup presented in [3] and formulate a 2D global grid approximation of the environment (see map M in Fig. 1) for storing visited positions, current location, and successful target recoveries. This concept is inspired by occupancy grid maps [7, 34], as opposed to post-exploration maps [32] or topological maps [20]. For our cattle recovery task – and despite their simplicity – grid maps still represent a highly effective tool [3] for exploring the solution space of AI solutions [49, 30]. Coat pattern identification of individual Friesian cattle represents a form of animal biometrics [23]. Early systems for the particular task at hand utilised the Scale-Invariant Feature Transform (SIFT) [26] on image sequences [27] or Affine SIFT (ASIFT) [33] to map from dorsal cow patterns to animal IDs [5]. However, for up-to-date performance we base our individual ID component on recent CNN-grounded biometric work [2] where temporal stacks of the regions of interest (RoIs) around detected cattle are analysed by a Long-term Recurrent Convolutional Network (LRCN) [10] as shown in Figure 1 in red. This architecture represents a compromise between light-weight onboard operation and more high-end networks with heavier computational footprints. Whilst aerial wildlife census applications routinely use manually controlled UAVs [16, 21, 1, 39] and have experimented with widely automated photographic gliders [31]. To the best of our knowledge this line of work presents the first proof-of-concept system for fully autonomous exploration and online individual biometric identification of animals completely onboard an aircraft.

Acknowledgements: the workshop presentation of this extended abstract was made possible by funding from the Artificial Intelligence grant from the Alan Turing Institute led by Professor Andrew Dowsey.

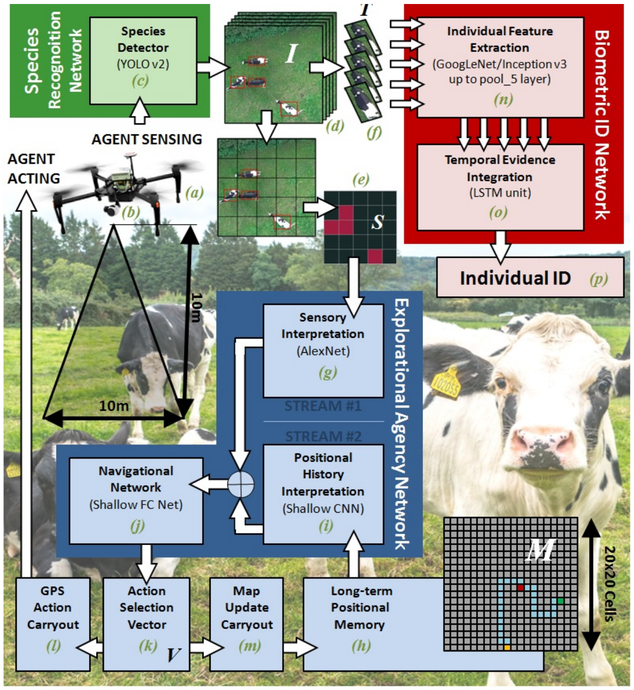

Figure 1. **Onboard System Overview.** For individual Friesian cattle search and identification we propose an approach that utilises three deep convolutional neural architectures operating onboard a ***(a)*** computationally-enhanced DJI Matrice 100 platform to achieve species detection (green), exploratory agency (blue), and individual animal ID (red). In particular, a ***(b)*** DJI Zenmuse X3 camera stream reshaped to 720×720 images is processed by ***(c)*** a YoloV2-based Friesian cattle detector yielding ***(d)*** RoI-annotated frames I. These are transformed into an ***(e)*** 5×5 occupancy grid map S and also a set of ***(f)*** spatio-temporal tracklets $\{T_0, T_1, ..., T_p\}$. The map S is input to a ***(g)*** tactical network based on AlexNet, which forms a dual stream navigation architecture together with a ***(i)*** strategic network operating on ***(h)*** a long-term explorational history memory M. Both streams are concatenated into ***(j)*** a shallow navigation net that outputs ***(k)*** a score vector V based on which possible navigational actions $a \in \{N, W, S, E\}$ are selected. During operation selected actions are ***(l)*** performed grounded in GPS and, in-turn, ***(m)*** the positional history M is updated. For individual ID ***(e)*** each tracklet T is rescaled and fed into an ***(n)*** Inception V3 network up to the pool_5 layer followed by ***(o)*** a LSTM unit for temporal information integration mapping to ***(p)*** a vector over the individual cattle IDs.

2. Hardware

We use the DJI Matrice 100 quadrotor UAV, which houses the ROS-enabled DJI N1 flight controller, as our base platform. It has been employed previously across various autonomous tasks [17, 8, 48, 24]. We extend the base M100 platform by adding an Nvidia Jetson TX2 mounted on a Connect Tech Inc. Orbitty carrier board to enable onboard deep inference via 256 CUDA cores under Nvidia's Pascal™ architecture. Also onboard is the DJI Manifold (essentially an Nvidia Jetson TK1) to decode the raw image feed from the onboard camera (the DJI Zenmuse X3 camera/gimbal) and to add further computational capacity. The X3 camera is mounted on rubber grommets and a 3-axis gimbal, which allows for rotor vibration isolation and independent movement of the flight platform for stabilised footage and programmatically controlled roll-pitch-yaw. A Quanum QM12V5AUBEC voltage regulation device was fitted to power non-conforming devices feeding off the primary aircraft battery. In addition, customised mounts were added for WiFi antennas to monitor the craft remotely (see Fig. 2).

3. Experimental Setup

Flights were performed over a two-week experimental period at the University of Bristol's Wyndhurst Farm in Langford Village, UK on a consistent herd of 17 yearling heifer Holstein Friesian cattle satisfying all relevant animal welfare and flight regulations. Experiments consisted of two phases: *(a)* training data acquisition across 14 (semi)manual flights, and subsequent *(b)* conduction of 18 autonomous flights. Training data was acquired over two day-long recording sessions where manual and semi-autonomous flights at varying altitudes were carried out, recording video at a resolution of 3840×2160 at 30fps. The result was a raw dataset consisting of 37 minutes from 15 videos over 14 flights occupying 18GB. Overall 2285 frames were extracted from these video files at a rate of $1Hz$. First, after discarding frames without cattle, 3120 bounding boxes around individual cattle were labelled in the 553 frames containing cattle. Animals were also manually identified as ground truth for individual identification. Secondly, to produce ground truth for training the cattle detector, square sub-images were manually annotated to encompass individuals such that they are resolved at approximately 150×150 pixels. Fig. 3 illustrates the full pipeline.

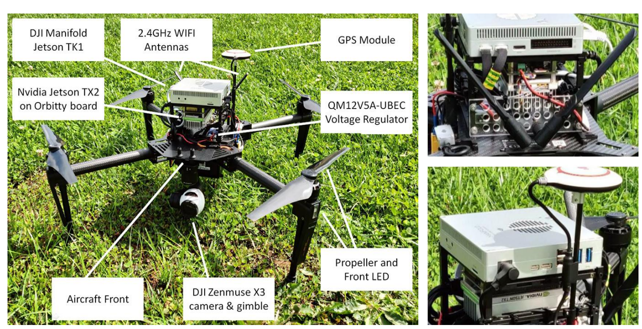

Figure 2. **Physical UAV Platform.** ***(left)*** Front view of the customised and fully assembled DJI Matrice 100 UAV flight platform with selected individual components highlighted. ***(right)*** Close-ups of the rear and side of the aircraft revealing custom mounts for two centrally fixed onboard computers, WiFi antennas and the GPS unit.

4. Implementation

Cattle detection and localisation is performed in realtime frame-by-frame using the YOLOv2 [38] CNN. The network was retrained from scratch on the annotated region dataset, consisting of $11,384$ original and synthetic training images and associated ground truth labels. Synthetic instances were generated via image augmentation; stochastic combinations of rotation, cropping, affine transformations, etc. Model inference operates on 736×736 images obtained by cropping and scaling the source 960×720 pixel camera stream. As shown in Figure 1, this process yields a set of m bounding boxes $B = \{b_0, b_1, ..., b_m\}$ per frame with associated object confidence scores. Inference on each of $n = 5$ sampled frames then produces a box-annotated spatio-temporal volume $\{B_0, B_1, ..., B_n\}$. Bounding boxes are associated across this volume by accumulating detections that are consistently present in cow-sized areas of an equally subdivided image. This method is effective for reliable short-term tracking due to distinct and non-overlapping targets, slow target movement and stable UAV hovering. The outputs are $p \leq n$ short individual animal tracklets reshaped into an image patch sequence $\{T_0, T_1, ..., T_p\}$, which forms the input to the individual identification network. In addition, the current frame I is also abstracted to a 5×5 grid map S encoding animal presence in the field of view of the camera (see Fig. 1).

Navigation activities aim at locating as many – themselves moving – individual animals as possible on the shortest routes in a gridded domain where a target counts as 'located' once the agent occupies the same grid location as the target. To solve this dynamic travelling salesman task (with city locations to be discovered on the fly), we use a dual-stream deep network architecture, as first suggested in our previous work [3]. The method computes grid based navigational decisions $\{N, W, S, E\}$ based on immediate sensory (tactical/exploitation) and historic navigational (strategic/exploration) information using two separate streams within a single deep inference network. As shown in the paper, this strategy can significantly outperform simple strategies such as a 'lawnmower' pattern and other baselines. To summarise the method's operation, the sensory input S is processed via a first stream utilising a basic AlexNet [22] design (see Fig. 1) because of its visual simplicity. A second stream operates on the exploratory history thus far, as stored in a long-term memory map M. This stores the agent's present and past positions alongside animal encounters within the flight area of 20×20 grid locations. The agent's starting position is fixed and M is reset after $\delta\%$ of the map has been explored. Both these streams are concatenated into a shallow integration network that, as shown in Figure 1, maps to a SoftMax normalised likelihood vector V of the possible navigational actions. During inference, the network selects the top-ranking navigational action from $\{N, W, S, E\}$ based on V, which is performed and, in-turn, the positional history M is updated. For training, the entire two-stream navigation network is optimised via stochastic gradient decent (SGD) with momentum [36] and a fixed learning rate $e = 0.001$ based on triples $(S; M; V)$ using one-hot encoding of V and cross-entropy loss. This unified model allows for the backpropagation of navigation decision errors across both streams and domains. For training, we simulate $10,000$ episodes of 17 pseudo-randomly [28] placed targets in a 20×20 grid and calculate optimal navigation decisions $(S; M; V)$ by solving the associated travelling salesman problem. 10-fold cross validation on this setup yielded an accuracy of 72.45% in making an optimal next grid navigation decision and a target recovery rate of 0.26 ± 0.06 targets per move.

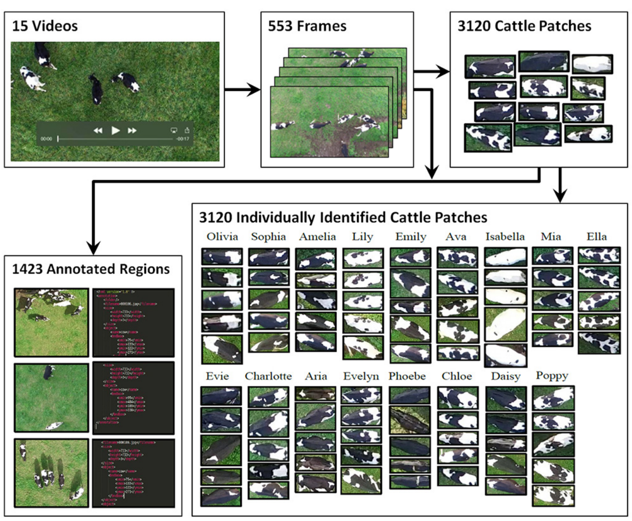

Figure 3. **Training Data Annotation.** Across all 15 videos gathered for training data acquisition, 553 frames contained cattle. Labelling bounding boxes around individual animals yielded 3120 cattle patches. Labelling 720×720 regions containing annotated cattle patches yielded the 1423 images annotated with cattle bounding boxes that form the base data for species detector training. Manual identification of the 17 cows, on the other hand, produced the base data for training the individual identification components.

Individual identification based on an image patch sequence $\{T_0, T_1, ..., T_p\}$ is performed via an LRCN, first introduced by Donahue et al. [10]. In particular, as shown in Figure 1, we combine a GoogLeNet/Inception V3 CNN [44, 45] with a single Long Short-Term Memory (LSTM) [15] layer. This approach has demonstrated success in disambiguating fine-grained categories in our previous work [2]. Training of the GoogLeNet/Inception V3 network takes groups of $n = 5$ same class randomly selected RoIs, each of which were non-proportionally resized to 224×224 pixels. SGD with momentum [36], a batch size of 32 and a fixed learning rate $e = 0.001$ were used for optimisation. Once

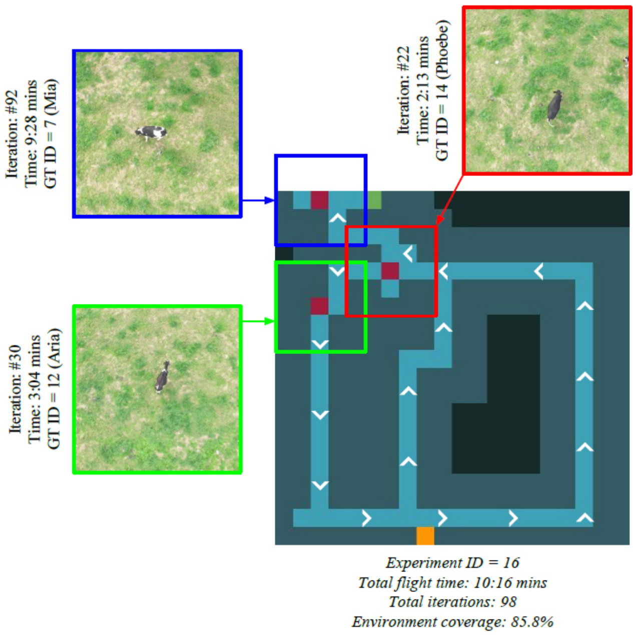

Figure 4. **Annotated Autonomous Flights.** Example of an annotated agent flight path within the exploratory grid over the entire course of the experiment. Cell colours are defined as (black): unvisited locations, (light blue): visited locations, (dark blue): seen or covered areas, (orange): agent starting position, (green): finishing agent position and (red): discovered target positions.

trained, samples are passed through this GoogLeNet up to the "pool 5 layer" and feature vectors are combined over the n samples. A shallow LSTM network is finally trained on these vector sequences using a SoftMax cross-entropy cost function optimised against the one-hot encoded identities vector representing the 17 possible classes.

5. Results

We conducted 18 fully autonomous test flights at a low altitude (approximately 10m) above an area of 20×20 cells covering altogether 147 minutes. For all experiments, we ran the object detection and exploratory agency networks live. Note that the herd was naturally dispersed in these experiments and animals were free to roam across the entire field in and out of the covered area. Thus, only a few individuals were present and recoverable in this area at any one time. The median coverage of the grid was 70.13% with median flight time of approximately 8 minutes per experiment, and a median of 77 grid iterations per flight.

For offline analysis, the UAV saved to file one acquired 720×720 image at each exploratory agency iteration, yielding 1039 images. 99 of those images actually contained target cows that were hand labelled with ground truth bounding box annotations and identities according to the VOC guidelines [11]. A detection was deemed a successful true positive based on the IoU ($ov \geq 0.5$). Grounded in this, we measured the YOLOv2 detection accuracy to be 92.4%, where out of the 111 present animals, 2 were missed and 7 false positive nested detections occurred. We then tested separately, the performance of the single frame Inception V3 individual identification architecture (yielding 93.6% accuracy), where all ground truth bounding boxes (not only the detected instances) were presented to the ID component. In contrast, and as shown in Table 1, when identification is performed on detected RoIs only then the combined offline system accuracy is 91.9%.

# Sample Frames	# Animal Instances	Detection on Sample Frames Accuracy (%)	Single Frame ID on Labelled RoIs Accuracy (%)	Combined Detection+ID Accuracy (%)
1039	111	92.4	93.6	91.9

Table 1. **Offline Performance Results.**

For online autonomous operation, all computation was performed live in real-time onboard the UAV's computers (DJI Manifold Nvidia Jetson TX2). Figure 4 depicts various example flights with detailed annotations of flight paths, animal encounters and identification confidences. Across the 1039 grid locations visited during the set of experiments, the aircraft navigated centrally above a detected individual 18 times, triggering identification. These individuals represented a small subset (6 individuals) of the larger herd that tended to be more comfortable with the presence of the aircraft directly above them. Note that this mode of operation eliminates the problem of clipped visibility at image borders, minimises image distortions, optimises the viewpoint, and exposes the coat in an orthogonal view. For triggered identification, we store LRCN confidence outputs after processing up to 5 same class patches T to compare performance differences between single view and multi-view identification. The respective overall results are given in Table 2. Notably, across the small sample set, the model performed perfectly.

# Samples	LRCN Identification Accuracy (%)	Single Frame Identification Accuracy (%)
18	**100**	94.4

Table 2. **Online Identification Results.**

6. Conclusion

The experiments conducted confirm demonstrable identification robustness of the approach. This gives rise to future agricultural automation possibilities with potential implications for animal welfare and farm productivity. Beyond farming, the concept of autonomous biometric animal identification from the air as presented opens up a realm of future applications in fields such as ecology, where animal identification of uniquely patterned species in the wild (e.g. zebras, giraffes) is critical to assessing the status of populations. Future work will investigate the effect of varying the UAV altitude and consequently, varied environment visibility, target resolution, etc.

References

[1] A. Abd-Elrahman, L. Pearlstine, and F. Percival. Development of pattern recognition algorithm for automatic bird detection from unmanned aerial vehicle imagery. *Surveying and Land Information Science*, 65(1):37, 2005.

[2] W. Andrew, C. Greatwood, and T. Burghardt. Visual localisation and individual identification of holstein friesian cattle via deep learning. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 2850–2859, 2017.

[3] W. Andrew, C. Greatwood, and T. Burghardt. Deep learning for exploration and recovery of uncharted and dynamic targets from uav-like vision. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1124–1131. IEEE, 2018.

[4] W. Andrew, C. Greatwood, and T. Burghardt. Aerial animal biometrics: Individual friesian cattle recovery and visual identification via an autonomous uav with onboard deep inference. *arXiv preprint arXiv:1907.05310*, 2019.

[5] W. Andrew, S. Hannuna, N. Campbell, and T. Burghardt. Automatic individual holstein friesian cattle identification via selective local coat pattern matching in rgb-d imagery. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 484–488. IEEE, 2016.

[6] F. Bonin-Font, A. Ortiz, and G. Oliver. Visual navigation for mobile robots: A survey. *Journal of intelligent and robotic systems*, 53(3):263, 2008.

[7] J. Borenstein and Y. Koren. Real-time obstacle avoidance for fast mobile robots in cluttered environments. In *Proceedings., IEEE International Conference on Robotics and Automation*, pages 572–577. IEEE, 1990.

[8] A. F. Cobo and F. C. Benıtez. Approach for autonomous landing on moving platforms based on computer vision. 2016.

[9] G. N. DeSouza and A. C. Kak. Vision for mobile robot navigation: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 24(2):237–267, 2002.

[10] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2625–2634, 2015.

[11] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

[12] A. Giusti, J. Guzzi, D. C. Cireşan, F.-L. He, J. P. Rodríguez, F. Fontana, M. Faessler, C. Forster, J. Schmidhuber, G. Di Caro, et al. A machine learning approach to visual perception of forest trails for mobile robots. *IEEE Robotics and Automation Letters*, 1(2):661–667, 2015.

[13] P. Gregorini. Diurnal grazing pattern: its physiological basis and strategic management. *Animal Production Science*, 52(7):416–430, 2012.

[14] P. Gregorini, S. Tamminga, and S. Gunter. Behavior and daily grazing patterns of cattle. *The Professional Animal Scientist*, 22(3):201–209, 2006.

[15] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[16] A. Hodgson, N. Kelly, and D. Peel. Unmanned aerial vehicles (uavs) for surveying marine fauna: a dugong case study. *PloS one*, 8(11):e79556, 2013.

[17] X. Hui, J. Bian, X. Zhao, and M. Tan. Vision-based autonomous navigation approach for unmanned aerial vehicle transmission-line inspection. *International Journal of Advanced Robotic Systems*, 15(1):1729881417752821, 2018.

[18] S. D. Jones, C. Andresen, and J. L. Crowley. Appearance based process for visual navigation. In *Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robot and Systems. Innovative Robotics for Real-World Applications. IROS'97*, volume 2, pages 551–557. IEEE, 1997.

[19] S. Kondo and J. F. Hurnik. Stabilization of social hierarchy in dairy cows. *Applied Animal Behaviour Science*, 27(4):287–297, 1990.

[20] A. Kosaka and A. C. Kak. Fast vision-guided mobile robot navigation using model-based reasoning and prediction of uncertainties. *CVGIP: Image understanding*, 56(3):271–329, 1992.

[21] W. R. Koski, T. Allen, D. Ireland, G. Buck, P. R. Smith, A. M. Macrander, M. A. Halick, C. Rushing, D. J. Sliwa, and T. L. McDonald. Evaluation of an unmanned airborne system for monitoring marine mammals. *Aquatic Mammals*, 35(3):347, 2009.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[23] H. S. Kühl and T. Burghardt. Animal biometrics: quantifying and detecting phenotypic appearance. *Trends in ecology & evolution*, 28(7):432–441, 2013.

[24] S. Kyristsis, A. Antonopoulos, T. Chanialakis, E. Stefanakis, C. Linardos, A. Tripolitsiotis, and P. Partsinevelos. Towards autonomous modular uav missions: The detection, geo-location and landing paradigm. *Sensors*, 16(11):1844, 2016.

[25] L. Lin and M. A. Goodrich. Uav intelligent path planning for wilderness search and rescue. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 709–714. IEEE, 2009.

[26] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 1150–1157. Ieee, 1999.

[27] C. A. Martinez-Ortiz, R. M. Everson, and T. Mottram. Video tracking of dairy cows for assessing mobility scores. 2013.

[28] M. Matsumoto and T. Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 8(1):3–30, 1998.

[29] Y. Matsumoto, M. Inaba, and H. Inoue. Visual navigation using view-sequenced route representation. In *Proceedings of IEEE International conference on Robotics and Automation*, volume 1, pages 83–88. IEEE, 1996.

[30] A. A. Melnikov, A. Makmal, and H. J. Briegel. Projective simulation applied to the grid-world and the mountain-car problem. *arXiv preprint arXiv:1405.5459*, 2014.

[31] MIT. Sloopflyer. https://caos.mit.edu/blog/glider-photography-sloopflyer. [Online; accessed 1-Mar-2019. Unpublished elsewhere].

[32] H. P. Moravec. The stanford cart and the cmu rover. *Proceedings of the IEEE*, 71(7):872–884, 1983.

[33] J.-M. Morel and G. Yu. Asift: A new framework for fully affine invariant image comparison. *SIAM journal on imaging sciences*, 2(2):438–469, 2009.

[34] G. Oriolo, M. Vendittelli, and G. Ulivi. On-line map building and navigation for autonomous mobile robots. In *Proceedings of 1995 IEEE International Conference on Robotics and Automation*, volume 3, pages 2900–2906. IEEE, 1995.

[35] C. Phillips and M. Rind. The effects of social dominance on the production and behavior of grazing dairy cows offered forage supplements. *Journal of Dairy Science*, 85(1):51–59, 2002.

[36] N. Qian. On the momentum term in gradient descent learning algorithms. *Neural networks*, 12(1):145–151, 1999.

[37] L. Ran, Y. Zhang, Q. Zhang, and T. Yang. Convolutional neural network-based robot navigation using uncalibrated spherical images. *Sensors*, 17(6):1341, 2017.

[38] J. Redmon and A. Farhadi. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263–7271, 2017.

[39] A. Rodríguez, J. J. Negro, M. Mulero, C. Rodríguez, J. Hernández-Pliego, and J. Bustamante. The eye in the sky: combined use of unmanned aerial systems and gps data loggers for ecological research and conservation of small birds. *PLoS One*, 7(12), 2012.

[40] K. J. Ryu. *Autonomous robotic strategies for urban search and rescue*. PhD thesis, Virginia Tech, 2012.

[41] P. Saeedi, P. D. Lawrence, and D. G. Lowe. Vision-based 3-d trajectory tracking for unknown environments. *IEEE transactions on robotics*, 22(1):119–136, 2006.

[42] J. Santos-Victor, G. Sandini, F. Curotto, and S. Garibaldi. Divergent stereo for robot navigation: Learning from bees. In *Proceedings of IEEE conference on computer vision and pattern recognition*, pages 434–439. IEEE, 1993.

[43] B. Sowell, J. Mosley, and J. Bowman. Social behavior of grazing beef cattle: Implications for management. In *Proceedings of the American Society of Animal Science*, volume 78, pages 1–6. Department of Animal and Range Sciences, Montana State University, Bozeman . . . , 1999.

[44] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[45] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[46] R. Ungerfeld, C. Cajarville, M. Rosas, and J. Repetto. Time budget differences of high-and low-social rank grazing dairy cows. *New Zealand journal of agricultural research*, 57(2):122–127, 2014.

[47] S. Waharte and N. Trigoni. Supporting search and rescue operations with uavs. In *2010 International Conference on Emerging Security Technologies*, pages 142–147. IEEE, 2010.

[48] H. Yu, S. Lin, J. Wang, K. Fu, and W. Yang. An intelligent unmanned aircraft system for wilderness search and rescue.

[49] J. Zhang, L. Tai, J. Boedecker, W. Burgard, and M. Liu. Neural slam: Learning to explore with external memory. *arXiv preprint arXiv:1706.09520*, 2017.